

END2END-ПОДХОД В ЗАДАЧАХ АВТОМАТИЧЕСКОГО РАСПОЗНОВАНИЯ РЕЧИ

Михайличенко И.В.

Научный руководитель – к.т.н., доц. Рахлис Д.Е.

Харьковский национальный университет радиоэлектроники
(61166, Харьков, пр. Науки, 14, каф. АПВТ, тел. (057) 702-13-26)

e-mail: ihor.mykhailichenko@nure.ua

Improving the quality and speed of speech recognition in noisy environments by creating a stable intellectual model, based on the application of the end2end approach, which makes it possible to minimize the loss of audio signals.

Введение. Идеологически система распознавания речи состоит из двух частей. Эти части могут быть неявно выделены в самостоятельные блоки или подпрограммы. Обе части могут существовать в упрощенном виде, но в любой реализации обе части присутствуют. Система распознавания речи состоит из акустической и лингвистической частей. В общем случае она может включать в себя фонетическую, фонологическую, морфологическую, лексическую, синтаксическую и семантическую модели языка. Или, наоборот, представлять собой простенький коррелятор. Акустическая модель отвечает за представление речевого сигнала. Вернее, за его преобразование (из традиционного временного процесса) в некоторую форму, в которой в более явном виде присутствует информация в содержании речевого сообщения [3]. Лингвистическая модель интерпретирует информацию, получаемую от акустической модели, и отвечает за представление результата распознавания потребителю, в роли которого может выступать не только человек, но и техническая система, управляемая речью. Цель исследования – повышение качества и быстродействия распознавания речи в условиях зашумленности за счёт создания устойчивой интеллектуальной модели, на основе применения end2end-подхода, что дает возможность минимизировать потери звуковых сигналов. Задача – разработка устойчивой модели взаимодействия человек-компьютер посредством голосовых команд.

Содержание исследования. End2End подход – это система, которая предназначена для того, чтобы напрямую отражать последовательность акустических признаков в последовательности графем (букв) или слов. Также можно сказать, что это система, которая оптимизирует критерии, напрямую влияющие на финальную метрику оценки качества [2]. Метод необходим для того, чтобы тренировать акустическую модель без необходимости пофреймового выравнивания между звуком и транскрипцией. Есть обычный распознаватель, который принимает на вход акустические признаки – выдает некие скрытые состояния, на основе

которых получаем условные вероятности. Распознаватель обычно из себя представляет несколько слоев нейронной сети. Стоит отметить, что метод оперирует помимо обычных символов еще специальным символом, который называется пустой символ или blank-символ [1]. Для того чтобы решить проблему, которая возникает из-за того, что не каждый акустический фрейм имеет фрейм в транскрипции и наоборот (то есть у нас есть буквы или звуки, которые звучат намного дольше, и есть короткие звуки, повторяющиеся звуки), и существует этот blank-символ.

Так же подход предназначен для того, чтобы максимизировать итоговую вероятность последовательностей символов и обобщать возможное выравнивание.

Выводы. Научная новизна определяется системной интеграцией end2end-подхода для распознавания человеческой речи в интеллектуальных системах, интеграция блока распознавателя, дополнительного blank-символа, что дает возможность качественного распознавания в условиях шума.

Список использованных источников:

1. Колбан Н. Распознавание речи в условиях шума. Методы решения / Н.Колбан. – 2010. – № 10. – С. 45-32.
2. Schwartz M. Internet of Things with ESP8266 / Marco Schwartz // Packt Publishing. – Birmigham-Mumbai. – 2016. – 226 с.
3. Микрофонный усилитель с АРУ и малошумящим смещением микрофона [Электронный ресурс] / Справочник по электронным компонентам. – Режим доступа: <http://www.gaw.ru/html.cgi/txt/ic/Maxim/amplifiers/voice/MAX9814.htm>. – Дата доступа: 15.02.20. – Загол. с экрана.